1971

Bayesian Statistics: The Fundamental Theorem

Carolyn Rhodes
Ouachita Baptist University

BAYESIAN STATISTICS (THE FUNDAMENTAL THEOREM)

Presented to

Dr. D. M. Seward

for

Honors Special Studies

H493

by

Carolyn Rhodes

Fall, 1971

BAYESIAN STATISTICS (THE FUNDAMENTAL THEOREM)

A. Introduction

B. Prior distribution

C. Posterior distribution

D. Expression and example of the rule

1. $$R_\alpha = \frac{k_\alpha \cdot w_\alpha^m (1-w_\alpha)^{n-m}}{\sum k_\alpha \cdot w_\alpha^m (1-w_\alpha)^{n-m}} \qquad (\alpha = 1, 2, 3, \ldots)$$

2. $$q^i = \frac{r_i \cdot s^i}{\sum_i r_i' \cdot s_i}$$

3. $$(B^i, A) = \frac{(B_i)(A, B_i)}{(B_1)(A, B_1) + (B_2)(A, B_2) + \ldots + (B_N)(A, B_N)}$$

E. Bing's paradox

F. Conclusion

G. Table of forms

H. Bibliography

BAYESIAN STATISTICS (THE FUNDAMENTAL THEOREM)

The problem of the foundation of statistics is to state a set of principles which entail the validity of all correct statistical inference, and which do not imply that any fallacious inference is valid. This sentence describes the purpose of much writing on statistical inference, in general, and Bayesian statistics, in particular.[1] Bayesian statistics was first introduced in a publication by Thomas Bayes in The London Philosophical Transactions, volumes 53 and 54 for the years 1763 and 1764, after Bayes' death in 1761.[2] However, since the entire statistical research of Bayes' involves enormous study, this paper will limit itself to the development and application of Bayes' fundamental theorem.

The starting point for a Bayesian analysis is the specification of a prior distribution for the unknown parameter. There is little argument about using a prior which is based on relative frequencies of past events. If one had records of the mean length of items produced each day by an industrial machine, most statisticians would agree that one should utilize this information in making an inference about the next day's production. However, disagreement arises when one supposedly has no information on which to base his prior. Now one must decide how to proceed. The answer is

---

[1]Donald L. Meyer, "Bayesian Statistics," Review of Educational Research, 36 (December, 1966), 503.

[2]Florian Cajori, A History of Mathematics (New York: The Macmillan Company, 1919), p. 230.

clear when it is recognized that a probability is a number associated with a degree of reasonable confidence and has no purpose except to give it a formal expression. If no information is relevant to the actual value of a parameter, the probability must be chosen so as to express the fact that no information is available. It must say nothing about the value of the parameter, except the bare fact that it may possibly be restricted to lie within certain definite limits.

Another approach is to restrict the prior to a class of "natural conjugate Bayes densities" (NCBD). An NCBD is a distribution such that if it is used as a prior, then the posterior density from Baye's theorem after observing a sample is another member of the same class. An example of an NCBD is the normal distribution for an unknown mean with known variance when the sample is from a normal distribution. If the prior for μ is normal (m, V), and a sample from normal (μ, σ^2) is observed, then the posterior distribution for μ is also normal (m', V'). The mean of the posterior distribution is a weighted average of the mean of the prior and the mean of the sample, where the weights are proportional to the reciprocals of the respective variances.

$$m' = \frac{\left(\bar{X}\frac{N}{\sigma^2} + \frac{m}{V}\right)}{\left(\frac{N}{\sigma^2} + \frac{1}{V}\right)}$$

Suppose $V = \frac{\sigma^2}{n'}$, allowing only even fractional values of n'. Choosing a prior variance is equivalent to choosing a "prior" sample size, n'. The expression of indifference may simply be a question of determining some kind of base point on a scale of information accumulation. In the above example, if $n' = 0$ is selected, the prior variance is infinity, and since the normal distribution approaches the uniform distribution as the variance approaches infinity, the indifference prior would be uniform. The denominator of m^1 is also the reciprocal of the posterior variance and the notation is equal to $(n+n')(\frac{1}{\sigma^2})$. With $N=O$, the posterior distribution is normal $(\bar{x}, \frac{\sigma^2}{n})$. If a symmetrical posterior probability interval for μ were constructed by adding and substitute 1.96 $\frac{\sigma^2}{n}$, then the Bayesian would say that his probability is 0.95 that μ lies between the calculated limits. Of course, the 0.95 confidence interval turns out to be exactly the same interval. The difference is that the non-Bayesian is incorrect if he interprets this interval in the probability sense above.[3]

The posterior distribution of the unknown parameter is the goal of a Bayesian analysis. Once this is attained, posterior probability intervals can be constructed, means and variances reported, and hypotheses regarding values of the parameters can be assigned a posterior degree of belief by integrating over the relevant subspace of the parameters. Since the prior distribution adopted for a posterior probability could be almost any form, the task of catagoring posterior distribution depending on prios even for standard sampling and experimental designs is virtually impossible.[4]

Following the discussion or prior distributions and posterior distributions, must come the proof and statement of the fundamental theorem. However, much has been written concerning the theorem and it would be impossible to include all the discussions, so only three were chosen. At first it may seen the discussions are irrelevant but bear in mind that all theorems have information leading to the proof of that theorem.

---

[3]Meyer, op. cit., p. 507.

[4]Ibid., p. 508.

I.

Let E denote a certain state or condition, which can appear under only one of the mutually exclusive complexes of causes: $F_1, F_2, \ldots$ and not otherwise. Let the probability for the actual existence of F_1 be k_1, and if F_1 really exists then let ω_1 be the productive probability for bringing forth the observed event, E (E being of a different nature from F), which can only occur after the previous existence of one of the mutually exclusive complexes, F. Let, in the same manner, F_2 have an existence probability of k_2 and a productive probability of ω_2, etc. If now, by actual observation, the event E has occurred exactly m times in n trials, then the probability that the complex F_1 was the origin of E is:

$$Q_1 = \frac{k_1 \cdot \omega_1^m (1-\omega_1)^{n-m}}{\sum k_\alpha \cdot \omega_\alpha^m (1-\omega_\alpha)^{n-m}} \qquad (\alpha = 1, 2, 3, \ldots)$$

Similarly that complex F_2 was the origin:

$$Q_2 = \frac{k_2 \cdot \omega_2^m (1-\omega_2)^{n-m}}{\sum k_\alpha \cdot \omega_\alpha^m (1-\omega_\alpha)^{n-m}} \qquad (\alpha = 1, 2, 3, \ldots)$$

and so on for the other complexes.

To prove this equation, let the number of equally possible cases in the general domain of action, which leads to one of the complexes F_α, be t. Furthermore, of these t cases let f_1 be favorable for the existence of complex F_1, f_2 for F_2, f_3 for F_3, ..., etc. Then the probabilities for the existence of the different complexes F_α $(\alpha = 1, 2, 3, \ldots)$ are:

$$k_1 = \frac{f_1}{t}; \quad k_2 = \frac{f_2}{t}; \quad k_3 = \frac{f_3}{t}; \ldots$$

Of the f_1 favorable cases for complex F_1, λ_1 are also favorable for the occurrence of E, f_2 favorable cases for complex F_2, λ_2 for the occurrence of E, etc. The probability of the happening E under the assumption that F_1 exists. The relative probability is $P_{F_1}(E)$ is: $\omega_1 = \frac{\lambda_1}{f_1}$ or in general

$$\omega_\alpha = \frac{\lambda_\alpha}{f_\alpha} \qquad (\alpha = 1, 2, 3, \ldots).$$

The total number of equally likely cases for the simultaneous occurrence of the event E with either one of the favorable cases for $F_1, F_2, F_3, \ldots$, is:

$$\lambda_1 + \lambda_2 + \lambda_3 + \ldots = \sum \lambda_\alpha$$

The number of favorable cases for the simultaneous occurrence of F_1 and E is λ_1, etc. Hence, we have measures for their corresponding probabilities.

$$Q_1 = \frac{\lambda_1}{\sum \lambda_\alpha} \qquad Q_2 = \frac{\lambda_2}{\sum \lambda_\alpha} \qquad \ldots \text{ but}$$

$$\lambda_1 = \omega_1 \cdot f_1 \qquad \lambda_2 = \omega_2 \cdot f_2 \qquad \ldots \text{ and}$$

$$f_1 = k_1 \cdot t \qquad f_2 = k_2 \cdot t \qquad \ldots \text{ hence}$$

$$\lambda_1 = \omega_1 \cdot k_1 \cdot t \qquad \lambda_2 = \omega_2 \cdot k_2 \cdot t \qquad \ldots$$

Substituting these values in the above expression for $Q_1, Q_2, \ldots$ then

$$Q_1 = \frac{k_1 \cdot \omega_1}{\sum k_\alpha \cdot \omega_\alpha} \qquad Q_2 = \frac{k_2 \cdot \omega_2}{\sum k_\alpha \cdot \omega_\alpha} \qquad \ldots$$

as the respective probabilities that the observed event originated from the complexes $F_1, F_2, F_3, \ldots$ Such probabilities are posterior probabilities.

Now, investigate the above expression for $Q_1, Q_2, \ldots$ The numerator in the expression for Q_1 is $k_1 \cdot \omega_1$, but k_1 is simply the prior productive probability of bringing forth the event observed from complex F_1. The product $k_1 \cdot \omega_1$ is simply the relative probability $P_{F_1}(E)$ or the probability that the event E originated from F_1. In the denominator, $\sum k_\alpha \omega_\alpha$ $(\alpha = 1, 2, 3, \ldots)$ is the total probability to get E from any of the complexes F_α. From this, the probability to get E exactly m times from F_1 in n total trials is:

$$P_1 = \binom{N}{m} k_1 \cdot \omega_1^m (1 - \omega_1)^{n-m}$$

and the probability to get E from any one of the complexes, F, m times out of n is: $\Sigma P_\alpha = \binom{N}{m} \Sigma k_\alpha \cdot w_\alpha^m (1-w_\alpha)^{n-m} \quad (\alpha = 1, 2, 3, \ldots).$

If by actual observation, E is to have happened exactly m times out of n, then the posterior probability that F was the origin is:

$$Q_1 = \frac{\binom{N}{m} k_1 \cdot w_1^m (1-w_1)^{n-m}}{\Sigma \binom{N}{m} k_\alpha \cdot w_\alpha^m (1-w_\alpha)^{n-m}} \quad (\alpha = 1, 2, 3, \ldots).$$

The factorials $\binom{N}{m}$ in numerator and denominator cancel each other. It is not assumed that the posterior probability is proportional to the prior probability.

Sometimes the different complex F may be of such special characters that their prior probabilities of existence are equal, $k_1 = k_2 = k_3 = \ldots = k_n$. In this case, the equation simply reduces to $Q_1 = \frac{w_1^m (1-w_1)^{n-m}}{\Sigma w_\alpha^m (1-w_\alpha)^{n-m}}$.

This equation gives the most general expression of the fundamental theorem which may be stated as follows:

If a definite observed event, E, can originate from a certain series of mutually exclusive complexes, F, and if the actual occurrence of the event has been observed, then the probability that it originated from a specified complex or a specified group of complexes is also the posterior probability or probability of existence of the specified complex or group of complexes.[5]

It happens frequently that the knowledge of the general domain of action is so incomplete that it is not possible to determine, a prior, the probability of the occurrence of a certain expected event. As stated earlier, this is nearly always the case with problems wherein organic life

---

[5] Fisher, Anne, The Mathematical Theory of Probabilities (New York: The Macmillan Company, 1922), 59.

enters as a determining factor or momentum, but the same state of affairs may also occur in the category of problems relating to games of chance. Suppose one had an urn which was known to contain white and black balls only, but the actual ration in which the balls of the two different colors were mixed, was unknown. With this knowledge beforehand, it is not possible to determine the probability for the drawing of a white ball. If, on the other hand, from actual experience, the results of former drawings from the same urn when the conditions in the general domain of action remained unchanged during each separate drawing, then these results might be used in the determination of the probability of a specified event by future drawings.

The problem may be stated in the most general form as follows: Let F_α denote a certain state or condition in the general domain of action, which state or condition can appear only in one or the other of the mutually exclusive forms, F_1, F_2, F_3... and not otherwise. Let the probability of existence of F_1, F_2, F_3... be k_1, k_2, k_3... respectively, and when one of the complexes F_1, F_2, F_3... exists let W_1, W_2, W_3... be the respective productive probabilities of bringing forth a specified event, E. If now, by actual observation, E happens m times out of n total trials, what is then the probability that the event E, will happen in the $(n+1)$ trial also?

By Bayes' Rule, the posterior probability or the probabilities of existence of the complexes F_1, F_2,... is:

$$Q_1 = \frac{k_1 \cdot \omega_1^m (1-\omega_1)^{n-m}}{\sum k_\alpha \cdot \omega_\alpha^m (1-\omega_\alpha)^{n-m}} \qquad Q_2 = \frac{k_2 \cdot \omega_2^m (1-\omega_2)^{n-m}}{\sum k_\alpha \cdot \omega_\alpha^m (1-\omega_\alpha)^{n-m}} \quad \ldots \quad (\alpha = 1, 2, 3, \ldots)$$

In the $(n+1)$th trial E may happen from any one of the mutually exclusive complexes F_1, F_2, F_3,... whose respective probabilities in producing the event, E, are W_1, W_2, W_3,... The addition theorem then gives the total probability of the occurrence of E in the $(n+1)$th trial is:

$$R_\alpha = \sum P_{F_\alpha}(E) = Q_1 \cdot w_1 + Q_2 \cdot w_2 + \cdots$$
$$= \frac{\sum k_\alpha \cdot w_\alpha^m (1-w_\alpha)^{n-m} \cdot w_\alpha}{\sum k_\alpha \cdot w_\alpha^m (1-w_\alpha)^{n-m}} \quad (\alpha = 1, 2, 3, \ldots)$$

If the prior probabilities of existence are of equal magnitude, the factors k in the above expression cancel each other in numerator and denominator and thus

$$R = \frac{\sum w_\alpha^m (1-w_\alpha)^{n-m} w_\alpha}{\sum w_\alpha^m (1-w_\alpha)^{n-m}} \quad (\alpha = 1, 2, 3, \ldots).$$

Example 1: An urn contains five balls of which a part is known to be white and the rest black. A ball is drawn four times in succession and replaced after each draw. By three of such drawings, a white ball was obtained and by one drawing a black ball. What is the probability that a white ball will be drawn in the fifth drawing?

In regard to the contents of the urn the following four hypotheses are possible:

F : 4 white, 1 black balls

F : 3 white, 2 black balls

F : 2 white, 3 black balls

F : 1 white, 4 black balls

Since nothing is known about the ratio of distribution of the different colored balls, by direct application of the principle of insufficient reason, the four complexes are regarded equally probable or: $k_1 = k_2 = k_3 = k_4 = 1/4$.

If either F_1, F_2, F_3, or F_4 exists, the respective productive probabilities are: $w_1 = 4/5$, $w_2 = 3/5$, $w_3 = 2/5$, $w_4 = 1/5$.

By a direct substitution in the formula:

$$R = \frac{\sum w_\alpha^m (1-w_\alpha)^{n-m} \cdot w_\alpha}{\sum w_\alpha^m (1-w_\alpha)^{n-m}} \quad (\alpha = 1, 2, 3 \ldots) \text{ for } n = 4 \text{ and } m = 3.$$

then:
$$R = \frac{(\frac{4}{5})^3(\frac{1}{5})(\frac{4}{5}) + (\frac{3}{5})^3(\frac{2}{5})(\frac{3}{5}) + (\frac{2}{5})^3(\frac{3}{5})(\frac{2}{5}) + (\frac{1}{5})^3(\frac{4}{5})(\frac{1}{5})}{(\frac{4}{5})^3(\frac{1}{5}) + (\frac{3}{5})^3(\frac{2}{5}) + (\frac{2}{5})^3(\frac{3}{5}) + (\frac{1}{5})^3(\frac{4}{5})} = \frac{47}{73}.$$ [6]

Bayes' Rule has been reduced to the most general form:

$$R = \frac{k_\alpha \cdot w_\alpha^m (1 - w_\alpha)^{n-m}}{\Sigma k_\alpha \cdot w_\alpha^m (1 - w_\alpha)^{n-m}} \qquad (\alpha = 1, 2, 3, \ldots)$$

This is an exact expression for the rule, but it is at the same time almost impossible to employ it in practice. Only in a few exceptional cases is a prior known, the different values of the often numerous probabilities of existence k_α, of the complexes F_α, and in order to apply the rule with exact results sufficient facts are required about the different complexes of causes from which the observed event, E, originated. Bayes deduced the rule from special examples resulting from drawing balls of different color from an urn where the different complexes of causes were materially existent. The probability of a cause or a certain complex of causes did not here mean the probability of existence of such a complex but the probability that the observed event originated from this particular complex. In order to elucidate this statement the following example is given.

Example II: Start with the following four hypotheses:

F : 4 white, 1 black balls

F : 3 white, 2 black balls

F : 2 white, 3 black balls

F : 1 white, 4 black balls

assigning 1/4 as the hypothetical existence probability.

By marking the five balls similarly as in the last example, with the numbers from 1 to 5, the following complexes are found:

F : 4 white, 1 black balls in $\binom{5}{1}$ ways

---

[6] Fisher, op. cit., p. 65.

F_2 : 3 white, 2 black balls in $\binom{5}{2}$ ways

F_3 : 2 white, 3 black balls in $\binom{5}{2}$ ways

F_4 : 1 white, 4 black balls in $\binom{5}{1}$ ways

This gives a total of 5 + 10 + 10 + 5 = 30 different complexes. Assuming all of these complexes equally likely to occur, the following probabilities of existence and productive probabilities exist. $k_1 = k_2 = \ldots = k_{30} = 1/30$

$w_1 = w_2 = w_3 = w_4 = w_5 = 4/5$ (productive probability for F_1)

$w_6 = w_7 = \ldots = w_{15} = 3/5$ (productive probability for F_2)

$w_{16} = w_{17} = \ldots = w_{25} = 2/5$ (productive probability for F_3)

$w_{26} = w_{27} = \ldots = w_{30} = 1/5$ (productive probability for F_4)

The total probability of getting a white ball in the second drawing is now

$$Q = \frac{w_\alpha^3 (1 - w_\alpha) w_\alpha}{\sum w_\alpha^3 (1 - w_\alpha)} \qquad (\alpha = 1, 2, 3, \ldots, 30),$$

Actual substitution of the above values of W in this formula gives the final result as : $Q = 17/28$ (see proof I) [7]

II.

This second discussion of Bayes' fundamental theorem has almost the same background information as the first. However, the equation representing the theorem is somewhat different. If the information sounds the same keep in mind that the theorems are the same only in different forms.

When an event has happened which may have been due to any one of a number of different causes, the question arises as to which cause has most probably been in action. It is possible, from an observed happening of the event, to draw any conclusions as to the relative probability of the various causes that may have led to it.

---

[7] Fisher, op. cit., p. 66.

P_{A_iB}/P_B is the probability that condition A_i is satisfied when condition B is known to be satisfied. Suppose that $A_1, A_2, \ldots, A_N$ are n conditions of which one must be satisfied, and only one can be satisfied when a trial is made. Then $P_B = \sum_i P_{A_iB}$

$$\frac{P_{A_iB}}{P_B} = \frac{P_{A_i}\,\frac{P_{A_iB}}{P_{A_i}}}{\sum_i P_{A_i}\,\frac{P_{A_iB}}{P_{A_i}}} = \frac{P_{A_i}\,P_{A_iB}}{\sum_i P_{A_i}\,P_{A_iB}} .$$

Suppose now that the event E may have any one of n distinct causes, of which in a given trial only one can come into play. Let condition B be that the event E shall happen, and condition A_i be that the i^{th} cause has come into play. Then P_{A_i} is the probability before the trial, then the i^{th} cause of E will come into play: P_{A_iB} is the probability that E will happen as a result of the i^{th} cause; and P_{A_iB}/P_B is the probability when E has happened, that it has happened as the result of the i^{th} cause. The formula may be conveniently written

$$Q_i = \frac{r_i \cdot S_i}{\sum_i r_i \cdot S_i}$$

Where r_i is the probability of the i^{th} cause before the result is known (the prior probability of the i^{th} cause); S_i is the probability of the event when the i^{th} cause is in action; and Q_i is the probability of the i^{th} cause, when the event is known to have happened (the posterior probability).

This is the fundamental theorem of Bayes' and so long as the r's and S's are known, there can be no ambiquity in applying it. The hesitation that is undoubtedly felt in making use of Bayes' formula depends upon the fact that, thought the S's are generally known, some assumption has to be made with respect to the r's, and the calculated probabilities of cause depend on the particular assumption made.

<u>Example 3</u>: A box contains n objects, each of which is either white or black, and each is equally likely to be drawn. An object is drawn and found to be

white. It is returned and an object is drawn again. What is the probability that it will be white?

Denote by P_r the prior probability that r of the objects are white. Then $rP_r/\sum rP_r$ is the posterior probability that r are white, and the probability of drawing a white object at the second trial is

$$\sum P_r r^2 / N \sum P_r r.$$

If it is assumed that P_r is independent of r and therefore equal to $\frac{1}{N}$, this last probability is $2/3 + 1/3N$. If, however, each object in the box is assumed to be equally likely black or white, then $P_r = \frac{N!}{r!(N-r)!} \cdot \frac{1}{2^N}$ and the required probability is $1/2 + 1/2^N$.

Example 4: A box contains a number N of objects not greater than m; and it is known that n of them are marked. It is assumed that when a set of m objects is drawn from a box, all sets of m are equally likely. A set of m is drawn and it is found that m of them are marked. What is the probability value of N?

It follows, from the data, that N is equal to or greater than N + m − m_1. The ~~probability~~ of the ~~observed~~ event, when the box contains N objects is

$$\frac{N!/m_1!(N-m_1)! \cdot (N-n)!/(m-m_1)!(N-n-m+m_1)!}{N!/m!(N-m)!}$$

that is

$$\frac{m!\,n!}{m_1!(n-m_1)!(m-m_1)!} \cdot \frac{(N-n)!(N-m)!}{N!(N-n-m+m_1)!}.$$

Hence if P_n is the prior probability that the box contains N objects, then, after the event, the probability is

$$q_N = \frac{f(N)P_n}{\sum_{n+m-m_1} f(N)P_n} \quad \text{where}$$

$$f(N) = \frac{(N-m)!(N-n)!}{N!(N-n-m+m_1)!}.$$

The most probable value of N is that which makes $f(N)\,P_n$ as great as possible.

Suppose first that all possible values of N are prior equally probable so that P_n is independent of N. Then the most probable value of N satisfied the inequalities

$$f(N) > f(N-1) \qquad\qquad f(N) > f(N+1)$$

which gives, for N, the greatest integer is nm/m_1.

Suppose next that $P_n \propto N$, so that large values of N are prior more likely than small ones. The most probable value of N is then given by $\quad Nf(N) > (N-1)f(N-1) \qquad\qquad Nf(N) > (N+1)f(N+1)$

which gives, for N, the greatest integer is $(n-1)(m-1)/(m_1-1)$.

If lastly $P_n \propto \frac{1}{N+1}$, so that small values of N are prior more likely than larger ones, the inequalities are

$$\frac{f(N)}{N+1} > \frac{f(N-1)}{N} \qquad\qquad \frac{f(N)}{N+1} > \frac{f(N+1)}{N+2}$$

giving for N, the greatest integer is $\{(m+1)(N+1)-m_1-2\}/(m_1+1)$.

Example 5: A and B play a game, at which A's prior chance of winning is equally likely to be $\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-2}{n}$ or $\frac{n-1}{n}$ Out of a set of a + b games, A wins a and loses b. What is the probable value of A's chance to win the next game?

If A's chances of winning is r/n, the probability of the observed result of the a + b games is $\quad \frac{(a+b)!}{a!\,b!}\left(\frac{r}{n}\right)^a\left(1-\frac{r}{n}\right)^b$.

Hence the posterior probable value of A's chances of winning the next game is

$$\frac{\sum_1^{N-1}\left(\frac{r}{n}\right)^{a+1}\left(1-\frac{r}{n}\right)^b}{\sum_1^{N-1}\left(\frac{r}{n}\right)^a\left(1-\frac{r}{n}\right)^b}.$$

Now, if n is not too small, the quantity $\frac{1}{n}\sum_1^{N-1}\left(\frac{r}{n}\right)^a\left(1-\frac{r}{n}\right)^b$ is very nearly equal to

$$\int_0^1 x^a(1-x)^b\,dx = \frac{a!\,b!}{(a+b+1)!}.$$

Hence, if n is large enough, the required result is very nearly equal to $\frac{a+1}{a+b+2}$. It has been assumed that the probability of A's chance of winning being measured by r/n is itself independent of r.

Suppose now that the probability of A's chance of winning being measured by r/n is proportional to $r/n(1-r/n)$ so that neither A nor B is extremely likely either to win or lose. Then the above expression, for the probable value of A's chance of winning the next game, becomes

$$\frac{\sum_1^{n-1}\left(\frac{r}{n}\right)^{a+2}\left(1-\frac{r}{n}\right)^{b+1}}{\sum_1^{n-1}\left(\frac{r}{n}\right)^{a+1}\left(1-\frac{r}{n}\right)^{b+1}}.$$

which is sensibly equal to $\frac{a+2}{a+b+4}$.[8]

Example 6: An observer watches the spinning of a coin, and notes the sequences of heads and tails. What is the probable number of spins that have occurred, when he has noted M sequences?

The number, N, of spins must be equal to or greater than M. On the supposition that the number of spins is N, the probability of the observed event is $\frac{(N-1)!}{(m-1)!(n-m)!}\cdot\frac{1}{2^{N-1}}$.

Hence, if the prior probability that the number of spins is N be represented by P_N, the probable number of spins is $\frac{\sum \frac{N!}{(N-m)!}\frac{1}{2^{N-1}}P_N}{\sum \frac{(N-1)!}{(N-m)!}\frac{1}{2^{N-1}}P_N}$.

On the assumption that all numbers of spins equal to or exceeding N are prior equally probable, that is

$$\frac{\sum_{N=m}^{\infty}\frac{N!}{(N-m)!}\cdot\frac{1}{2^{N-1}}}{\sum_{N=m}^{\infty}\frac{(N-1)!}{(N-m)!}\cdot\frac{1}{2^{N-1}}}$$

$$= \frac{m\left(1-\frac{1}{2}\right)^{-m-1}}{\left(1-\frac{1}{2}\right)^{-m}}$$

$$= 2m.$$

---

[8] William Burnside, Theory of Probability (London: Cambridge University Press, 1936), pp. 57-59.

Moreover, on the same assumption, the most probable value of N is 2M.

Now, in this question, it is not a reasonable assumption that all values of N above M are equally probable. The spinning must take time and for this reason there must be an upper limit to N. If it is assumed that all values of N from M to M are equally probable, the probable value of N is $m\,{}^{A}/_{B}$, where

$$A = 1 + \frac{m+1}{1}\cdot\frac{1}{2} + \frac{(m+1)(m+2)}{2!}\cdot\frac{1}{2^2} + \dots + \frac{(m+1)(m+2)\dots m'}{(m'-m)!}\cdot\frac{1}{2^{m'-m}}$$

$$B = 1 + \frac{m}{1}\cdot\frac{1}{2} + \frac{m(m+1)}{2!}\cdot\frac{1}{2^2} + \dots + \frac{m(m+1)\dots(m'-1)}{(m'-m)!}\cdot\frac{1}{2^{m'-m}}$$

so that

$$A-B = \frac{1}{2} + \frac{m+1}{1}\cdot\frac{1}{2^2} + \dots + \frac{(m+1)(m+2)\dots(m'-1)}{(m'-m-1)!}\cdot\frac{1}{2^{m'-m}}$$

$$= \tfrac{1}{2}\left\{A - \frac{(m+1)(m+2)\dots m'}{(m'-m)!}\cdot\frac{1}{2^{m'-m}}\right\}.$$

Hence the probable value of N is

$$2m\left\{1 - \frac{(m+1)(m+2)}{(m'-m)!}\cdot\frac{1}{2^{m'-m}}\cdot\frac{1}{2B}\right\}.$$

This is always less than 2M.

It has been seen above that when N is large the probable number of sequence in N spins is $\frac{1}{2}$N, the duration of the spins not affecting the question. When, however, a number of M sequences are observed, and the corresponding probable number of spins is to be determined, the question of duration does affect the question and the probable number of spins is less than 2M.

<u>Example 7</u>: There are M counters, marked from 1 to M, in a bag and one is drawn, each being equally likely to be taken. The counter marked N is drawn and a coin equally likely to fall head or tail is spun 2N times and the excess 2n of heads over tails is noted. This is repeated 5 times. 2N spins being made each time and the excesses of heads over tails are found to be $n_1, n_2, \dots, n_5$. The whole proceeding with the numbers

$m, n_1, n_2, n_3, n_4, n_5$ is reported to a calculator, the number N only being withheld from him. What conclusions can he draw about N?

The prior probability that N has any given value from 1 to M is $1/m$. If $|N_1|$ is the greatest of the positive numbers $|N_1|, |N_2|, \ldots, |N_S|$, the probability of the observed set of excesses of heads over tails is zero, when $N < (N_1)$ and is $\left\{ \frac{(2N)!}{2^{2N}} \right\}^S \prod_{i=1}^{i=S} \frac{1}{(N+n_i)!(N-n_i)!}$.

The approximate value of this latter expression is

$$\frac{1}{(\pi N)^{S/2}} e^{-\frac{1}{N}\sum_1^S Y_i^2}.$$

If then $N > (n_1)$ the calculator infers that the probability that the counter drawn was marked N is $\dfrac{1/N^{S/2}\, e^{-\sigma/N}}{\sum_{n_1}^{m} 1/N^{S/2}\, e^{-\sigma/N}}$ where $\sigma = \sum_1^S Y_1^2$.

The most probable value of N is that which makes $e^{-\sigma/N}/N^{S/2}$ as great as possible. The maximum value of this quantity when N varies continuously, is given by $N = \frac{2\sigma}{S}$; so that the most probable value of N is one of the integers on either side of $2\sigma/S$.

Since $e^{-\sigma/N}/N^{S/2}$ when sensible in values, changes little when N is changed to N + 1, the probability that N lies between N and N may be written, approximately $\dfrac{\int_{N_1}^{N_2} N^{-S/2} e^{-\sigma/N} dN}{\int_0^{\infty} N^{-S/2} e^{-\sigma/N} dN}$.

Putting $\sigma = NX$, this is

$$\frac{1}{\Gamma(S/2)} \int_{\sigma/N_2}^{\sigma/N_1} X^{S/2 - 1} e^{-X} dX$$ [9]

9Burnside, op. cit., pp. 62-64.

III.

An event A can occur only if one of the set of exhaustive and incompatible events B_1, B_2, ...B_N occurs. The probabilities of these events (B_1), (B_2),...(B_N) corresponding to the total absence of any knowledge as to the occurrence or nonoccurrence of A are known. Known also, are the conditional probabilities

$$(A, B_i) \qquad i = 1, 2, \ldots, N$$

for A to occur assuming the occurrence of Bi. How does the probability of Bi change with the additional information that A has actually happened? The question amounts to finding the conditional probability (Bi, A). The probability of the compound event A Bi can be presented in two forms

$$(AB_i) = (B_i)(A, B_i)$$
$$(AB_i) = (A)(B_i, A) .$$

Equating the right-hand members, the following expression is derived for the unknown probability (Bi, A): $(B_i, A) = \dfrac{(B_i)(A, B_i)}{(A)}$.

Since the event A can materialize in the mutually exclusive forms $AB_i, AB_2, \ldots, AB_n$ by applying the theorem of total probability,

$$A = (B_1)(A, B_1) + (B_2)(A, B_2) + \ldots + (B_N)(A, B_N).$$

It suffices now to introduce this expression into the preceeding formula for (Bi, A) to get the final expression

$$(B_i, A) = \frac{(B_i)(A, B_i)}{(B_1)(A, B_1) + (B_2)(A, B_2) + \ldots + (B_N)(A, B_N)}$$

This formula is known as the formula for probabilities of hypotheses. The reason for that name is that the events B_1, B_2, ...B_N may be considered as hypotheses to account for the occurrence of A. It is customary to speak of probabilities $(B_1), (B_2), \ldots, (B_N)$
as prior probabilities of hypotheses $B_1, B_2, \ldots, B_N$
while probabilities $(B_i, A) \qquad i = 1, 2, 3, \ldots, N$

are called posterior probabilities of the same hypotheses.

Example 8: The contents of urns 1, 2, 3 are as follows:

1 white, 2 black, 3 red balls

2 white, 1 black, 1 red balls

4 white, 5 black, 3 red balls

One urn is chosen at random and two balls drawn. They happen to be white and red. What is the probability that they come from urn 2 or 3?

The event A represents the fact that two balls taken from the selected urn were of white and red color respectively. To account for this fact, there are 3 hypotheses. The selected urn was 1 or 2 or 3. These are represented by B_1, B_2, B_3. Since nothing distinguishes the urns, the probabilities of these hypotheses before anything was known about A are $(B_1) = (B_2) = (B_3) = 1/3$.

The probabilities of A, assuming these hypotheses are

$$(A, B_1) = 1/5 \qquad (A, B_2) = 1/3 \qquad (A, B_3) = 2/11 .$$

It now remains to introduce those values into the formula to have a posterior probabilities:

$$(B_2, A) = \frac{1/3 \cdot 1/3}{1/3 \cdot 1/5 + 1/3 \cdot 1/3 + 1/3 \cdot 2/11} = 55/118$$

$$(B_3, A) = \frac{1/3 \cdot 2/11}{1/3 \cdot 1/5 + 1/3 \cdot 1/3 + 1/3 \cdot 2/11} = 30/118$$

$$(B_1, A) \neq \frac{55 - 30}{118} = 33/118$$

Retaining the notations, conditions, and data of the problem, find the probability of materialization of another event C granted that A has actually occurred. Conditional probabilities (C, AB_i'); are supposed to be known.

Since the fact of the occurrence of A involves that of one, and only one, of the events, B_1, B_2, ... B_N, the event C can materialize in the following mutually exclusive forms CB_1, $CB_2, \ldots,$ CB_N. Consequently, the probability (C, A) which is being sought is given by

$$(C, A) = (CB_1, A) + (CB_2, A) + \ldots + (CB_N, A).$$

Applying the theorem of compound probability,

$$(CB_i, A) = (B_i, A)(C, B_i A)$$
$$(C, A) = (B_1, A)(C, AB_1) + (B_2, A)(C, AB_2) + \ldots + (B_N, A)(C, AB_N).$$

It suffices now to substitute for (B_i, A) its expression given by Bayes' formula to find the final expression

$$(C, A) = \frac{\sum_{i=1}^{N} (B_i)(A, B_i)(C, AB_i)}{\sum_{i=1}^{N} (B_i)(A, B_i)}.$$

It may happen that the materialization of hypotheses B_i makes C independent of A, then $(C, AB_i) = (C, B_i)$

and the equation reduces simply to

$$(C, A) = \frac{\sum_{i=1}^{N} (B_i)(A, B_i)(C, B_i)}{\sum_{i=1}^{N} (B_i)(A, B_i)}.$$ [10]

By making an extended use of the infinitesimal calculus, Mr. Bing and Dr. Kroman in their memoirs arrived at much more ambiguous results through an application of the rule of Bayes. Starting with the fundamental rule

$$Q = \frac{\binom{N}{m} k_1 \cdot w_1^m (1 - w_1)^{n-m}}{\sum \binom{N}{m} k_\alpha \cdot w_\alpha^m (1 - w_\alpha)^{n-m}} \qquad (\alpha = 1, 2, 3, \ldots)$$

simple conditions inside the domain of causes can be encountered. The total complex of actions may embrace a large number of smaller sub-complexes construed in such a way be regarded as a continuous process, so that the productive probabilities are increased by an infinitely small quantity from a certain lower limit, a, to an upper limit, b. Denoting such continuously increasing probabilities by V and the corresponding small

---

[10] J. V. Uspensky, Introduction to Mathematical Probability (New York: McGraw-Hill Book Company, 1937).

probabilities of existence by $u\,dv$, the total probability of obtaining E from any one of the minor complexes with a productive probability between α and β $(\alpha \geq a, \beta \leq b)$ is $P = \int_{\alpha}^{\beta} uv\,dv$. The probability that when E has happened it originated from one of these minor complexes, or the probability of existence of some one of those complexes is:

$$P = \frac{\int_{\alpha}^{\beta} uv\,dv}{\int_{a}^{b} uv\,dv}$$

The situation may be still more simplified by the following considerations. In the continuous total complex between the limits a and b is situated $(b-a)/dv$ individual minor complexes. Assuming all of these complexes to possess the same probability of existence, then $u\,dv = \frac{dv}{b-a}$

The two formulas then take on the form $P = \frac{1}{b-a}\int_{\alpha}^{\beta} v\,dv$

$$P = \frac{\int_{\alpha}^{\beta} v\,dv}{\int_{a}^{b} v\,dv}.$$

A still more specialized form is obtained by letting a = o and b = 1 which gives: $P = \int_{\alpha}^{\beta} v\,dv$

$$P = \frac{\int_{\alpha}^{\beta} v\,dv}{\int_{0}^{1} v\,dv}.$$

Example 9: An urn contains a very large number of similarly shaped balls. In ten successive drawings (and replacements), 7 with the number 1, 2 with the number 2, and 1 with the number 3 were obtained. What is the probability to obtain a ball with another number in the following drawing?

The balls are marked 1, 2, 3 or "others." A general scheme of distribution of the ball in the urn may be given through the following scheme:

nx balls marked with the number 1

ny balls marked with the number 2

nz balls marked with the number 3

nt = n(1-x-y-z) other balls.

Hence x, y, z, and t represent the respective productive probabilities. Let such probabilities assume all possible values between 0 and 1 with intervals of $1/N$, the possible conditions in the total complex of actions is obtained. Each of these conditions has a probability of existence, s, and the productive probability x, y, z, and 1-x-y-z. The original probability for 7 ones, 2 twos, and 1 three in ten drawings is

$$P = \frac{10!}{7!\,2!\,1!} \sum s \cdot x^7 \cdot y^2 \cdot z\,.$$

Now when n is a very large number, the interval $1/N$ becomes a very small quantity, and may approximately be written: $s = u\,dx\,dy\,dz$

and also write the above sum as a triple integral:

$$P = \frac{10!}{7!\,2!\,1!} \int_0^1 \int_0^p \int_0^q u \cdot x^7 \cdot y^2 \cdot z \; dx\,dy\,dz$$

where p=1-x and q=1-x-y. If now the above event has happened, then the probability to get a different marked ball in the eleventh drawing is:

$$Q = \frac{\int_0^1 \int_0^p \int_0^q u \cdot x^7 \cdot y^2 \cdot z(1-x-y-z)\,dx\,dy\,dz}{\int_0^1 \int_0^p \int_0^q u \cdot x^7 \cdot y^2 \cdot z\;dx\,dy\,dz}$$

It is quite impossible to evaluate the above integral without knowing the form of the function u, but unfortunately the information at hand tells absolutely nothing in regard to this. Perhaps the balls bear the numbers 1, 2, 3 only or perhaps there is an equal distribution up to 10,000 or any other number. The information is really so insufficient that it is quite hopeless to attempt a calculation of the posterior probability.

Many adherents of the inverse probability method venture boldly forth with the following solution based upon the perfectly arbitrary hypotheses: that all the u's are of equal magnitude. This gives the special integral

$$Q = \frac{\int_0^1 \int_0^p \int_0^q x^7 \cdot y^2 \cdot z(1-x-y-z)\,dx\,dy\,dz}{\int_0^1 \int_0^p \int_0^q x^7 \cdot y^2 \cdot z\;dx\,dy\,dz} \qquad x+y+z \leq 1.$$

In this case the limits of x are 0 and 1, those of y are 0 and 1-x, those of z are 0 and 1-x-y.

This is a well-known form of the triple integral which may be evaluated by means of Dirichlets' Theorem:

$$u = \int_0^1 \int_0^{1-x} \int_0^{1-x-y} x^{b-1} y^{m-1} z^{n-1} \, dx\,dy\,dz$$
$$= \frac{\Gamma(b)\,\Gamma(m)\,\Gamma(n)}{\Gamma(1+b+m+n)} .$$

Remembering the well-known relation between gamma functions and factorials, $\Gamma(n+1) = n!$ by a mere substition in the integral, the value of the probability in question is 1:14. Another and equally plausible result is obtained by a slightly different working of the problem.

The successive drawings have resulted in balls marked 1, 2, or 3. What is the probability to obtain a ball not bearing such a number in the eleventh drawing? This probability is given by the formula

$$\frac{\int_0^1 v^{10}(1-v)\,dv}{\int_0^1 v^{10}\,dv} = 1:12 .$$

Quite a different result from the one given above.[11]

A more astonishing paradox is produced by Bing when he gives an example of Bayes Rule to a problem from motality statics. A motality table gives the ratio of the number of persons living during a certain period, to the number living at the beginning of this period, all persons being of the same age. By recording the deaths during the specified period (one year) it has been ascertained that of s persons, say forty years of age at the beginning of the period, m have died during the period. The observed ratio is then (s-m)/s. If s is a very large number this ratio may be taken as an approximation of the true ratio of probability if survived

[11]Fisher, op. cit., pp. 70-72.

during this period. If s is not sufficiently large, the believers in the inverse theory ought to be able to evaluate this ratio by an application of Bayes' Rule, by means of an analysis similar to the one that follows:

Let y be the general symbol for the probability of a forty year old person being alive one year from hence. Each of such persons will in general be subject to different conditions, and the general symbol, y, will therefore have to be understood as the symbol for all the possible productive probability values changing from 0 to 1 by a continuing process.

Assuming s a very large number each condition will have a probability of existence equal to udy. What is the probability that the ratio of survival of a group of s persons aged forty is situated between the limits α and β?

The answer according to Bayes' Rule is:

$$\frac{\int_{\alpha}^{\beta} y^{s-m}(1-y)^{m}\, u\, dy}{\int_{0}^{1} y^{s-m}(1-y)^{m}\, u\, dy} \qquad I$$

Let us furthermore divide the whole year into two equal parts and let y_1 be the probability of surviving the first half year, y_2 the probability of surviving the second half year, and $u_1 dy_1$, $u_2 dy_2$ the corresponding probability of existence. Then the respective posterior probability of y_1 and y_2 are:

$$\frac{y_1^{s-m}(1-y_1)^{m_1}\, u_1\, dy_1}{\int_0^1 y_1^{s-m_1}(1-y_1)^{m_1}\, u_1\, dy_1}$$

$$\frac{y_2^{s-m}(1-y_2)^{m_2}\, u_2\, dy_2}{\int_0^1 y_2^{s-m}(1-y_2)^{m_2}\, u_2\, dy_2} \qquad (m_1+m_2=m)$$

(m_1 and m_2 represent the number of deaths in the respective half years). The probability that both y_1 and y_2 are true is then according to the multiplication theorem:

$$\frac{y_1^{s-m_1}(1-y_1)^{m_1}\, u_1\, dy_1 \;\; y_2^{s-m}(1-y_2)^{m_2}\, u_2\, dy_2}{\int_0^1 y_1^{s-m_1}(1-y_1)^{m_1}\, u_1\, dy_1 \int_0^1 y_2^{s-m}(1-y_2)^{m_2}\, u_2\, dy_2}$$

where $y = y_1 \cdot y_2$.

The probability that the probability of survival for a full year, y, is situated between the limits α and β is therefore:

$$\frac{\iint y_1^{s-m_1}(1-y_1)^{m_1} y_2^{s-m}(1-y_2)^{m_2} u_1 \cdot u_2 \cdot dy_1 \cdot dy_2}{\int_0^1 y_1^{s-m_1}(1-y_1)^{m_1} u_1 dy_1 \int_0^1 y_2^{s-m}(1-y_2)^{m_2} u_2 dy} \qquad II$$

where the limits in the double integral in the numerator are determined by the relation: $\alpha \leq y_1 y_2 \leq \beta$.

Choosing the principle of insufficient reason as the basis of calculations, merely assuming that all possible events are, in the absence of any grounds for inference, equally likely, the various quantities expressed by the general symbol, u, become equal and constant and cancel each other in numerator and denominator, which brings the posterior probability expressed by (I) and (II) to the forms:

$$\frac{\int_\alpha^\beta y^{s-m}(1-y)^m dy}{\int_0^1 y^{s-m}(1-y)^m dy} \qquad III$$

$$\frac{\iint y_1^{s-m_1}(1-y_1)^{m_1} y_2^{s-m_1-m_2}(1-y_2)^{m_2} dy_1 \cdot dy_2}{\int_0^1 y_1^{s-m_1}(1-y_1)^{m_1} \int_0^1 y_2^{s-m}(1-y_2)^{m_2} dy_1 \cdot dy_2} \qquad IV$$

where the limits in the numerator in the latter expression are determined by the relation: $\alpha < y_1 y_2 < \beta$. Letting $y_2 = y/y_1$ and then $1-y_1 = z(1-y)$ this latter expression may after a simple substitution be brought to the form:

$$\frac{\int_\alpha^\beta y^{s-m}(1-y)^{m+1} dy \int_0^1 \frac{z^{m_1}(1-z)^{m_2} dz}{1-z(1-y)}}{\int_{y_1} y_1^{s-m_1}(1-y_1)^{m_1} dy_1 \int_0^1 y_2^{s-m}(1-y_2)^{m_2} dy_2}.$$ [12]

(see Proof II). This will result in a different probability from the equation (I) found by Bayes' Rule, however, both used the same discussion and proof.

---

[12]Fisher, op. cit., pp. 73-74.

Whether or not the reader agrees with Bayes' theorem and the uses of this theorem, he must investigate all aspects of the theorem plus the discussion and background that led to it. Bayes, however, is assured of his immortality since he was the first to use mathematical probability inductively, "that is for arguing from the particular to the general, or from the sample to the population."[13]

---

[13] E. T. Bell, Development of Mathematics (New York: McGraw-Hill Book Company, Incorporated, 1945), p. 583.

TABLE OF FORMS

1. $$R_\alpha = \frac{k_\alpha \cdot w_\alpha^m (1-w_\alpha)^{n-m}}{\Sigma k_\alpha \cdot w_\alpha^m (1-w_\alpha)^{n-m}} \qquad (\alpha = 1, 2, 3, \ldots)$$

2. $$q_i = \frac{r_i' \cdot s_i'}{\sum_i r_i' \cdot s_i'}$$

3. $$(B_i, A) = \frac{(B_i)(A, B_i)}{(B_1)(A, B_1) + (B_2)(A, B_2) + \ldots + (B_N)(A, B_N)}$$

4. $$(C, A) = \frac{\sum_{i=1}^{N} (B_i)(A, B_i)(C, B_i)}{\sum_{i=1}^{N} (B_i)(A, B_i)}$$

5. $$Q = \sum_i P(\theta = \theta_i)\, \rho(\theta_i, d)$$ [14]

6. $$\sum_{\theta=1}^{k} b(\theta) r(\theta; s) \leq \sum_{\theta=1}^{k} b(\theta) r(\theta; t)$$ [15]

7. $$P(B_j' \backslash X_i) = \frac{P(B_j')\, P(X_i \backslash B_j')}{\sum_j P(B_j')[P(X_i \backslash B_j')]}$$ [16]

8. $$P(B_i \backslash A) = \frac{P(B_i \cap A)}{P(A)}$$ [17]

---

[14] B. W. Lindgren and G. W. McElrath, Introduction to Probability and Statistics (New York: The Macmillan Company, 1959).

[15] Lionel Weiss, Statistical Decision Theory (New York: McGraw-Hill Book Company, 1961).

[16] Samuel B. Richmond, Statistical Analysis (New York: Ronald Press Company, 1964).

[17] W. A. Thompson, Applied Probability (New York: Holt, Rinehart and Winston, Incorporated, 1969).

PROOF I

Numerator:

w_1, w_2, w_3, w_4, w_5 $\qquad Q_1 = \left(\frac{4}{5}\right)^3\left(\frac{1}{5}\right)\left(\frac{4}{5}\right)5 = \frac{256}{625}$

$w_6, w_7, \cdots, w_{15}$ $\qquad Q_2 = \left(\frac{3}{5}\right)^3\left(\frac{2}{5}\right)\left(\frac{3}{5}\right)10 = \frac{324}{625}$

$w_{16}, w_{17}, \ldots, w_{25}$ $\qquad Q_3 = \left(\frac{2}{5}\right)^3\left(\frac{3}{5}\right)\left(\frac{2}{5}\right)10 = \frac{96}{625}$

$w_{26}, w_{27}, \ldots, w_{30}$ $\qquad Q_4 = \left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)\left(\frac{1}{5}\right)5 = \frac{4}{625}$

Denominator:

w_1, w_2, w_3, w_4, w_5 $\qquad Q_5 = \left(\frac{4}{5}\right)^3\left(\frac{1}{5}\right)5 = \frac{64}{125}$

$w_6, w_7, \ldots, w_{15}$ $\qquad Q_6 = \left(\frac{3}{5}\right)^3\left(\frac{2}{5}\right)10 = \frac{108}{125}$

$w_{16}, w_{17}, \ldots, w_{25}$ $\qquad Q_7 = \left(\frac{2}{5}\right)^3\left(\frac{3}{5}\right)10 = \frac{48}{125}$

$w_{26}, w_{27}, \ldots w_{30}$ $\qquad Q_8 = \left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)5 = \frac{4}{125}$

$$R = \frac{Q_1 + Q_2 + Q_3 + Q_4}{Q_5 + Q_6 + Q_7 + Q_8}$$

$$= \frac{680}{625} \Big/ \frac{224}{125}$$

$$= \frac{136}{224}$$

$$= \frac{17}{28}$$

PROOF II

$$\frac{\iint y_1^{s-m_1}(1-y_1)^{m_1}\, y_2^{s-m_1-m_2}(1-y_2)^{m_2}\, dy_1 \cdot dy_2}{\int_0^1 y_1^{s-m_1}(1-y_1)^{m_1} \int_0^1 y_2^{s-m}(1-y_2)^{m_2}\, dy_1 \cdot dy_2} \qquad \text{IV}$$

The double integral is of the form $\iint_{(A)} F(y_1, y_2)\, dy_1\, dy_2$ where (A) is defined by means of relations: $\alpha < y_1 y_2 < \beta,\ 0 < y_1 < 1,\ 0 < y_2 < 1.$ The field of integration is thus the area swept out by the hyperbola $y_1 y_2 = \alpha$, the st. line $y_2 = 1$, the hyperbola $y_1 y_2 = \beta$ and the st. line $y_1 = 1$. Changing the variables by means of the transformation:

$$y_1 y_2 = y = \varphi(y, z) \qquad 1 - y_1 = z(1-y) = \psi(y, z)$$

we get the following new double integral

$$\iint_{(A_1)} = \left[\varphi(y,z),\ \psi(y,z)\right] /J/\, dy\, dz \qquad (/J/ \text{ absolute value})$$

where J is the Jacobian or functional determinant defined by the formula:

$$J = \begin{vmatrix} \frac{d\varphi}{\partial y} & \frac{d\varphi}{\partial z} \\ \frac{d\psi}{\partial y} & \frac{d\psi}{\partial z} \end{vmatrix} = \frac{d\varphi}{\partial y}\frac{\partial \psi}{\partial z} - \frac{\partial \varphi}{\partial z}\frac{\partial \psi}{\partial y} \quad \text{for}$$

$$\frac{y}{y_1} = y_2 = \frac{y}{1-z(1-y)} = \varphi(y,z) \qquad y_1 = 1 - z(1-y) = \psi(y,z)$$

$$|J| = \left\| \begin{matrix} \frac{1-z}{[1-z(1-y)]^2} & \frac{y(1-y)}{[1-z(1-y)]^2} \\ z & -(1-y) \end{matrix} \right\| = \frac{1-y}{1-z(1-y)}$$

The transformation in a double integral implies in general 3 parts (1) the expression of $F(y_1, y_2)$ in terms of y, z;(2) the determination of the new system of limits;(3) sub. of dy_1, dy_2. The third solved above. The solution of the two first is purely algebraically.

NUMERATOR:

$$\iint y_1^{s-m_1}(1-y_1)^{s-m_1-m_2}(1-y_2)\, dy_1 \cdot dy_2$$

$$m_1 + m_2 = m\ ;\ dy_1 \cdot dy_2 = \frac{1-y}{y_1}\, dy\, dz\ ;\ y_1 = (1-z(1-y)]\ ;\ y_2 = \frac{y}{y_1}$$

$$= \iint y_1^{s-m_1}(1-y_1)^{m_1}\left(\frac{y}{y_1}\right)^{s-m}\left(\frac{y_1-y}{y_1}\right)^{m_2}\left(\frac{1-y}{y_1}\right) dy\, dz$$

$$= \iint \frac{(1-y_1)^{m_1}(y^{s-m})(y_1-y)^{m_2}(1-y)\, dy\, dz}{y_1}$$

$$\iint \frac{[z(1-y)]^{m_1}(y_0^{s-m})(1-z(1-y)-y)^{m_2}(1-y)\,dy\,dz}{y_1} \qquad y_0 = Y$$

$$\iint \frac{[z(1-y)]^{m_1}(y^{s-m})[(1-y)(1-z)]^{m_2}(1-y)\,dy\,dz}{y_1}$$

$$\iint \frac{(y^{s-m})z^{m_1}(1-z)^{m_2}(1-y)^{m_1+m_2+1}\,dy\,dz}{y_1}$$

$$\iint \left[y^{s-m}(1-y)^{m+1}\right]\left[\frac{z^{m_1}(1-z)^{m_2}}{y_1}\right]dy\,dz$$

The easiest way to determine the new system of limits is probably by constructing the contour in the new field of integration. The hyperbolas $y_1 y_2 = \alpha$ and $y_1 y_2 = \beta$ are in the new field of integration changed into the two straight lines $y = \alpha$ and $y = \beta$ which determines the limits for the variable y. An inspection of the expressions for $\varphi(y, z)$ and $\psi(y, z)$ shows that the two straight lines $y_2 = 1$ and $y_1 = 1$ become in the new field $z = 1$ and $z = 0$ which are the limits of z. The contour (A_1) simply becomes a rectangle bounded by the straight lines $z = 0$, $y = \beta$, $z = 1$, and $y = \alpha$ The complete transformation finally becomes

$$\frac{\int_\alpha^\beta y^{s-m}(1-y)^{m+1}\,dy \int_0^1 \frac{z^{m_1}(1-z)^{m_2}\,dz}{1-z(1-y)}}{\int y_1^{s-m_1}(1-y_1)^{m_1}\,dy_1 \int_0^1 y_2^{s-m}(1-y_2)^{m_2}\,dy_2}$$

BIBLIOGRAPHY

Bell, E. T. Development of Mathematics. New York: McGraw-Hill Book Company, 1945.

Burnside, William. Theory of Probability. London: Cambridge University Press, 1936.

Cajori, Florian. A History of Mathematics. New York: The Macmillan Company, 1919.

Fisher, Anne. The Mathematical Theory of Probabilities. New York: The Macmillan Company, 1922.

Lindgren, B. W. and G. W. McElrath. Introduction to Probability and Statistics. New York: The Macmillan Company, 1959.

Meyer, Donald L. "Bayesian Statistics," Review of Educational Research, XXXVI (December, 1966), 503-516.

Richmond, Samuel B. Statistical Analysis. New York: Ronald Press Company, 1964.

Thompson, W. A. Applied Probability. New York: Holt, Rinehart, and Winston, Incorporated, 1969.

Uspensky, J. V. Introduction to Mathematical Probability. New York: McGraw-Hill Book Company, Incorporated, 1937.

Weiss, Lionel. Statistical Decision Theory. New York: McGraw-Hill Book Company, Incorporated, 1961.